Rec'd PCT/PTO 19 JUL 2004 #2

PCT/IB 03 / 00156
20. 01. 03

Europäisches Patentamt

European Patent Office

Office européen des brevets

REC'D 28 JAN 2003
WIPO PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterlagen stimmen mit der ursprünglich eingereichten Fassung der auf dem nächsten Blatt bezeichneten europäischen Patentanmeldung überein.

The attached documents are exact copies of the European patent application described on the following page, as originally filed.

Les documents fixés à cette attestation sont conformes à la version initialement déposée de la demande de brevet européen spécifiée à la page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02290155.7

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk

EPA/EPO/OEB Form 1014.1 - 02.2000 7001014

BEST AVAILABLE COPY

Anmeldung Nr:
Application no.: 02290155.7
Demande no:

Anmeldetag:
Date of filing: 22.01.02
Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

Koninklijke Philips Electronics N.V.
Groenewoudseweg 1
5621 BA Eindhoven
PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:
(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.
If no title is shown please refer to the description.
Si aucun titre n'est indiqué se referer à la description.)

Drift-free video encoding and decoding method

In Anspruch genommene Prioriät(en) / Priority(ies) claimed /Priorité(s) revendiquée(s)
Staat/Tag/Aktenzeichen/State/Date/File no./Päys/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

H04N7/26

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of filing/Etats contractants désignées lors du dépôt:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE TR

"DRIFT-FREE VIDEO ENCODING AND DECODING METHOD"

FIELD OF THE INVENTION

The present invention relates to an encoding method for the compression of an original video sequence divided into successive groups of frames (GOFs) and to a corresponding decoding method.

BACKGROUND OF THE INVENTION

Current video standards (from MPEG-1 to H.26 L) often use so-called hybrid solutions. An hybrid video encoder is based on a predictive scheme where each frame is temporally predicted from a given reference frame (the prediction options being, as shown in Fig.1 : zero value prediction, for the intra frames, or I frames, forward prediction, for the P frames, or bi-directional prediction, for the B frames). The prediction error is then spatially transformed to get advantage of spatial redundancies (a 2D-DCT transform is used in the standard schemes).

A different approach has been proposed in the document "Three-dimensional subband coding of video", C.Podilchuk and al., IEEE Transactions on Image Processing, vol.4, n°2, February 1995, pp.125-139. A Group of Frames (GOF) is processed as a three-dimensional (2D+t) structure and spatio-temporally filtered in order to compact the energy in the low frequencies (further studies included Motion Compensation in this scheme in order to improve the overall coding efficiency). The 3D subband structure is depicted in Fig.2. The well known SPIHT algorithm was extended from 2D to 3D in order to efficiently encode the final coefficient bit-planes with respect to the spatio-temporal decomposition structure.

As it is implemented now, a 3D subband codec applies the motion-compensated (MC) spatio-temporal analysis at the full original resolution. Spatial scalability is achieved by getting rid of the highest spatial subbands of the decomposition. When motion compensation is used in the 3D analysis scheme, this method does not allow a perfect reconstruction of the video sequence at lower resolution, even with an infinite bit-rate (this phenomena will be referred to as drift in the following description). As explained in the document "Multiscale video compression using wavelet transform and motion compensation", P.Y.Cheng and al., Proceedings of the International Conference on Image Processing (ICIP95), Vol.1, 1995, pp.606-609, this drift comes from the order of wavelet transform and motion compensation that is not interchangeable. Indeed, when a frame (A) is synthesized at a lower resolution (a), the following operation is applied :

$$\begin{aligned} a &= DWT_L(L) + MC[DWT_L(H)] \\ &= DWT_L(A) + [MC[DWT_L(H)] - DWT_L(MC[H])] \end{aligned} \qquad (1)$$

where DWT_L denotes the resolution downsample using the same wavelet filters as in the 3D analysis. In a perfect scalable solution, one wants to have:

$$a = DWT_L (A) \qquad (2)$$

The remaining part of the expression (1) therefore corresponds to the drift. It can be noticed that, if no MC is applied, the drift is removed. The same phenomena happens (except at the image borders) if a unique motion vector is applied to the frame. Yet, it is known that MC is unavoidable to achieve a good coding efficiency, and the likelihood of a unique global motion is small enough to eliminate this particular case in the following paragraphs.

Some authors, such as J.W.Woods and al in the document "A resolution and frame-rate scalable subband/wavelet video coder", IEEE Transactions on Circuits and Systems for Video Technology, vol.1, n°9, September 2001, pp.1035-1044, get rid of this drift to achieve good spatial scalability by different means. However, in said document, the described scheme, in addition to being quite complex, implies the sending of an extra information (the drift correction necessary to correctly synthesize the upper resolution) in the bitstream, thus wasting some bits (the solution described in the document "Multiscale video compression..." avoids this bottleneck but works on a predictive scheme and is not transposable to the 3D subband codec).

SUMMARY OF THE INVENTION

It is therefore an object of the invention to propose a solution avoiding these drawbacks.

To this end, the invention relates to a video encoding method for the compression of an original video sequence divided into successive groups of frames (GOFs), said method comprising the steps of :

(1) generating from the original video sequence, by means of a wavelet decomposition, a low resolution sequence including successive low resolution GOFs ;

(2) performing on said low resolution sequence a low resolution decomposition, by means of a motion compensated spatio-temporal analysis of each low resolution GOF ;

(3) generating from said low resolution decomposition a full resolution sequence, by means of an anchoring of the high frequency spatial subbands resulting from the wavelet decomposition to said low resolution decomposition ;

(4) coding said full resolution sequence and the motion vectors generated during the motion compensated spatio-temporal analysis, for generating an output coded bitstream.

The proposed solution is remarkable in the sense that the global structure of the decomposition tree in the 3DS analysis is preserved and no extra information is sent to correct the drift effect (only the decomposition/reconstruction mechanism is

changed). If no motion estimation/compensation is performed at full resolution, it is a low-cost solution in terms of complexity. If motion compensation is introduced in the high spatial subbands, a better coding efficiency is provided.

The invention also relates to a corresponding decoding method, comprising the steps of :

(1) decoding said input coded bitstream for generating a decoded full resolution sequence and associated decoded motion vectors ;

(2) in said decoded full resolution sequence, separating the decoded high frequency spatial subbands and the decoded low resolution decomposition ;

(3) generating from said decoded low resolution decomposition, by means of a motion compensated spatio-temporal synthesis, a decoded low resolution sequence ;

(4) reconstructing from said decoded low resolution sequence and the decoded high frequency spatial subbands an output full resolution sequence corresponding to the original video sequence.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in a more detailed manner, with reference to the accompanying drawings in which :

- Fig.1 illustrates the different predictions in a typical hybrid video encoding scheme ;
- Fig.2 shows a 3D subband decomposition ;
- Fig.3 depicts an embodiment of an encoding scheme according to the invention ;
- Fig.4 depicts an embodiment of a decoding scheme corresponding to the encoding scheme of Fig.3 ;
- Fig.5 illustrates the reordering of the high spatial subbands (for a forward motion compensation) ;
- Fig.6 depicts another embodiment of an encoding scheme according to the invention.

DETAILED DESCRIPTION OF THE INVENTION

The proposed solution (i.e. a spatial scalability with no drift in a motion compensated 3D subband codec) is now explained with reference to its two main steps : (a) motion compensation at the lowest resolution, (b) encoding the high spatial subbands.

First in order to avoid drift at lower resolutions, Motion Compensation (MC) is applied at this level. Consequently one first downsizes the GOF using the wavelet filters. Then the usual 3D subband MC-decomposition scheme is applied to this downsized GOF (it may be noticed that a side effect of this method is the reduction of

the amount of motion vectors to be sent in the bitstream, which saves up some bits for texture coding). Before transmitting the subbands to a tree-based entropy coder (for instance to a 3D-SPIHT encoder such as described for instance in the document "Low bit-rate scalable video coding with 3D set partitioning in hierarchical trees (3D-SPIHT)", B.J. Kim and al. IEEE Transactions on Circuits and Systems for Video Technology, vol.10, n°8, December 2000, pp.1374-1387, one puts the high spatial subbands that allow the reconstruction of the full resolution. The final tree structure looks very similar to that of a 3D subband codec such as the one described in the document "A fully scalable 3D subband video codec"', V.Bottreau and al. Proceeding of IEEE Conference on Image Processing (ICIP2001), vol.2, pp.1017-1020, Thessaloniki, Greece, October 7-10, 2001, and so a tree-based entropy coder can be applied on it without any restriction, as described in the new encoding scheme of Fig.3. The corresponding decoding scheme, depicted in Fig.4, is symmetric to this encoder. To enable spatial scalability, the high frequency spatial subbands just have to be cut as in the usual version of the 3DS codec, the decoding scheme of Fig. 4 showing how to naturally obtain the low resolution sequence.

Then, for coding the high spatial subbands, two main solutions are proposed, the first one without MC, and the second one with MC.

A) Without MC

In the first solution, the high subbands simply correspond to the high frequency spatial subbands of the original (full resolution) frames of the GOF in the wavelet decomposition. Those subbands allow the reconstruction at full resolution at the decoder. Indeed, the frames can be decoded at the low resolution. However, these frames correspond to the low spatial subband in the wavelet analysis of the original frames. Hence one has merely to put the low resolution frames and the corresponding high subbands together and apply a wavelet synthesis to obtain the full resolution frames. But now, where and how to put those high subbands in order to optimize the 3D-SPIHT encoder ? In a MC scheme for a 3D subband encoder, the low temporal subbands always look like one of the original frames of the GOF. As a matter of fact :

$$L = \frac{1}{\sqrt{2}} \ [A + MC(B)] \qquad (3)$$

so L looks like A. Consequently, the high spatial subband of A should be placed with the low resolution decomposition corresponding to L. This approach (reordering of the high spatial subband in the case of forward motion compensations) is illustrated in Fig.5, where DWT_H denotes the high frequency wavelet filter and the coefficients c_{jt} are multiplication coefficients. The way to define c_{jt} is described later.

However, the motion compensation in the 3D subband structure can be either forward or backward (it has even been shown that alternate directions improve coding efficiency. The following algorithm , in which the notations are :

- . jt : temporal decomposition level (0 for the full frame-rate, jt_max for the lowest framerate)
- . t : 0 for the Low temporal subband, 1 for the High one
- . nf : subband index at temporal level jt
- . me_dir_desc_tree : a byte that describes the ME directions used at a given temporal level jt (the LSB describes the direction of the first ME/MC, 0 means "forward", 1 means "backward"),

makes the link between a frame GOF_index in the GOF and the spatio-temporal subband {jt;n;t} which resembles it most, depending on the Motion Estimation Direction Description Tree.

```
UInt8
STlocationToGofIndex(MEDirectionDescriptionTree me_dir_desc_tree, UInt8
jt_max, UInt8 jt, UInt8  nf, UInt8 t)
{
    UInt8     gof_index=0 ;
    UInt8     direction ;
    UInt8     j,n_sb ;
    UInt8     sign ;

    gof_index = nf<<jt ;

    sign = 1 ;
    n_sb = nf ;

    for ( j=jt-1 ; j>=0 ; j--)
      {
         direction = 1<<n_sb ;
         if (t==0)
         sign=0 ;
         direction &= me_dir_desc_tree.aui8_level[j] ;
         direction >>= n_sb ;
         if (sign)
         {
             direction = !direction ;
             sign = 0 ;
```

```
        }
        n_sb = (n_sb<<1) + direction ;

        direction <<= j ;
        gof_index  = direction;
    }
    return(gof_index) ;
}
```

The way to define the coefficients c_{jt} is now described (in Haar filter case). Let α be the coefficient used in the temporal 2-tap Haar filter. In the conventional 3D subband scheme, one has :

$$\begin{cases} L = \alpha * (A + MC^{-1}(B) \\ H = \alpha * (MC(A) - B) \end{cases}$$

If, in the present scheme, one uses $c_{jt} = \alpha^{jt}$ for the high spatial subbands, then it is still meaningful to use temporal scalability. Indeed :

$$\begin{cases} DWT_L(L) = \alpha * (\ DWT_L(A) + MC^{-1}(DWT_L(B))\ ) \\ DWT_H(L) = c_{jt} * (\ DWT_H(A)\ ) \\ = \alpha * (\ DWT_H(A + UpSample[MC^{-1}(DWT_L(B))]\ ) \end{cases}$$

and :

$$\begin{cases} DWT_L(H) = \alpha * (\ DWT_L(B) - MC[DWT_L(A)]\ ) \\ DWT_H(H) = \alpha * (\ DWT_H(B)\ ) \end{cases}$$

where UpSample refers to the picture upsizing using wavelet filters. For the reconstruction at a lower frame rate, only the low temporal subband is synthesized :

$$\begin{cases} \hat{L} = \dfrac{1}{2 * \alpha} DWT^{-1}[DWT(L)] \\ \quad = \dfrac{1}{2} * (\ A + UpSample[MC^{-1}(DWT_L(B))]\ ) \end{cases}$$

Finally, the reconstructed frames at each temporal level will tend to look like a motion-compensated average of the "reference" original frame and a blurred version of the other one (up-sampled version of the downsized frame), whereas in the current version of the 3D subband codec this blur is not introduced. Improving spatial scalability at the expense of adding blur in the temporal scalability is however a worthy step.

B) With MC

As using MC in every subband does not allow a reconstruction with no drift, it is possible, as depicted in Fig.6, to partially use MC to construct the high spatial

subbands (which is better in terms of coding efficiency) and still be able to reconstruct every resolution. Instead of directly using the high frequency spatial subbands of the wavelet decomposition, a wavelet decomposition is carried out on a prediction error obtained from the MC performed on the full resolution sequence and reusing for instance the motion vectors of the low resolution.

The solution is to define :

$$\begin{cases} DWT_H(L) = c_{jt} * (DWT_H(A)) \\ DWT_H(H) = c_{jt} * DWT_H(B - MC(A)) \end{cases}$$

It can be noticed that the MC is only used in the high temporal subband : A is first reconstructed at the full resolution thanks to the low temporal subband, and then used to get frame B with MC thanks to H. The coefficients c_{jt} are chosen as previously. Said MC at full resolution can be performed either by merely upsampling the low resolution motion vectors (which has the advantage of introducing no other motion vector overhead) or by refining these upsampled low resolution vectors (which costs some additional transmission bits but is more efficient in terms of texture coding).

CLAIMS :

1. A video encoding method for the compression of an original video sequence divided into successive groups of frames (GOFs), said method comprising the steps of:

(1) generating from the original video sequence, by means of a wavelet decomposition, a low resolution sequence including successive low resolution GOFs ;

(2) performing on said low resolution sequence a low resolution decomposition, by means of a motion compensated spatio-temporal analysis of each low resolution GOF ;

(3) generating from said low resolution decomposition a full resolution sequence, by means of an anchoring of the high frequency spatial subbands resulting from the wavelet decomposition to said low resolution decomposition ;

(4) coding said full resolution sequence and the motion vectors generated during the motion compensated spatio-temporal analysis, for generating an output coded bitstream.

2. A method according to claim 1, in which, for each frame, said high spatial subbands are directly anchored to the low resolution subband that, in said spatio-temporal decomposition, looks most like said frame, depending on the motion estimation direction.

3. A method according to claim 1, in which a predictive mode is used to construct the high spatial subbands, said high spatial subbands resulting from a second wavelet decomposition performed on a prediction error obtained from a motion compensation applied to the original video sequence.

4. A method for decoding an input bitstream coded by means of an encoding method according to anyone of claims 1 to 3, said decoding method comprising the steps of :

(1) decoding said input coded bitstream for generating a decoded full resolution sequence and associated decoded motion vectors ;

(2) in said decoded full resolution sequence, separating the decoded high frequency spatial subbands and the decoded low resolution decomposition ;

(3) generating from said decoded low resolution decomposition, by means of a motion compensated spatio-temporal synthesis, a decoded low resolution sequence ;

(4) reconstructing from said decoded low resolution sequence and the decoded high frequency spatial subbands an output full resolution sequence corresponding to the original video sequence.

Abstract

The invention relates to a video encoding method for the compression of a video sequence, comprising the steps of generating from the original video sequence, by means of a wavelet decomposition, a low resolution sequence, performing on said low resolution sequence a low resolution decomposition, by means of a motion compensated spatio-temporal analysis, generating from said low resolution decomposition a full resolution sequence, by means of an anchoring of the high frequency spatial subbands resulting from the wavelet decomposition to said low resolution decomposition and coding said full resolution sequence and the motion vectors generated during the motion compensated spatio-temporal analysis. The invention also relates to a corresponding decoding method.

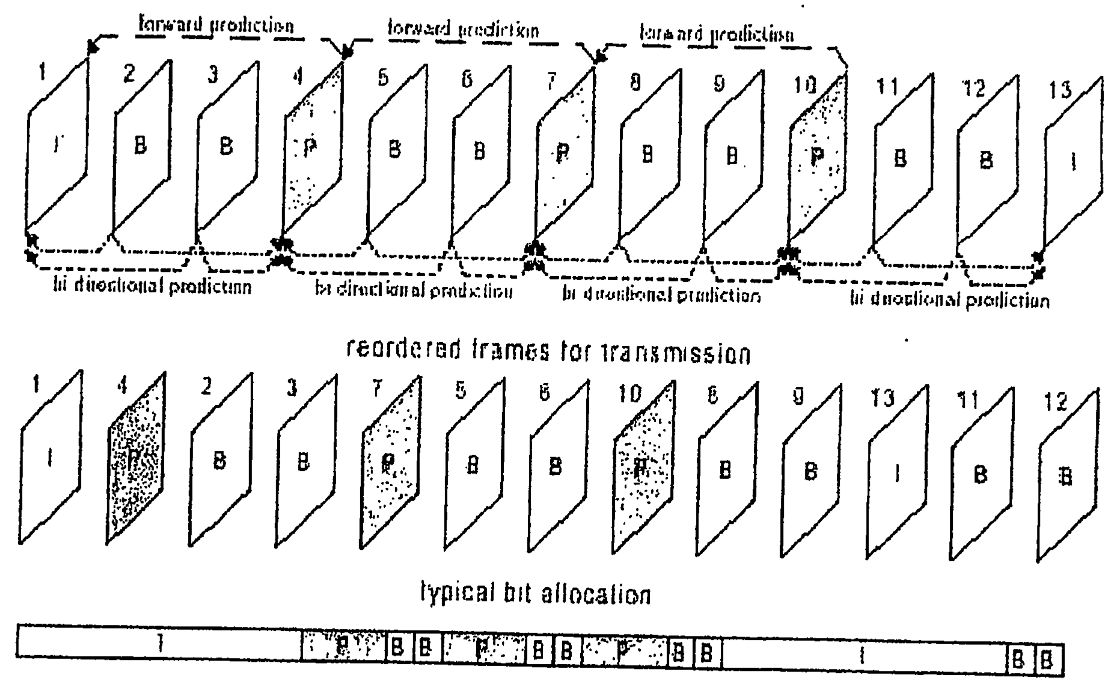

forward prediction
forward prediction
forward prediction
1 2 3 4 5 6 7 8 9 10 11 12 13
I B B P B B P B B P B B I
bi directional prediction
bi directional prediction
bi directional prediction
bi directional prediction
reordered frames for transmission
1 4 2 3 7 5 6 10 8 9 13 11 12
I P B B P B B P B B I B B
typical bit allocation
I P B B P B B P B B I B B

FIG.1

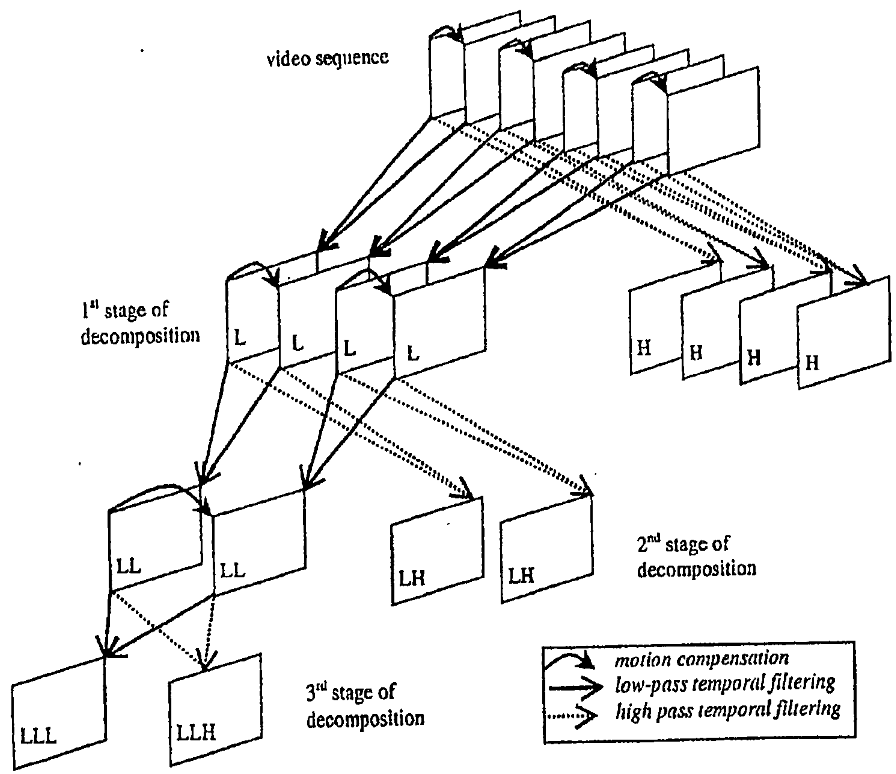

video sequence
1st stage of decomposition
L
L
L
L
H
H
H
H
LL
LL
LH
LH
2nd stage of decomposition
LLL
LLH
3rd stage of decomposition
motion compensation
low-pass temporal filtering
high pass temporal filtering

FIG. 2

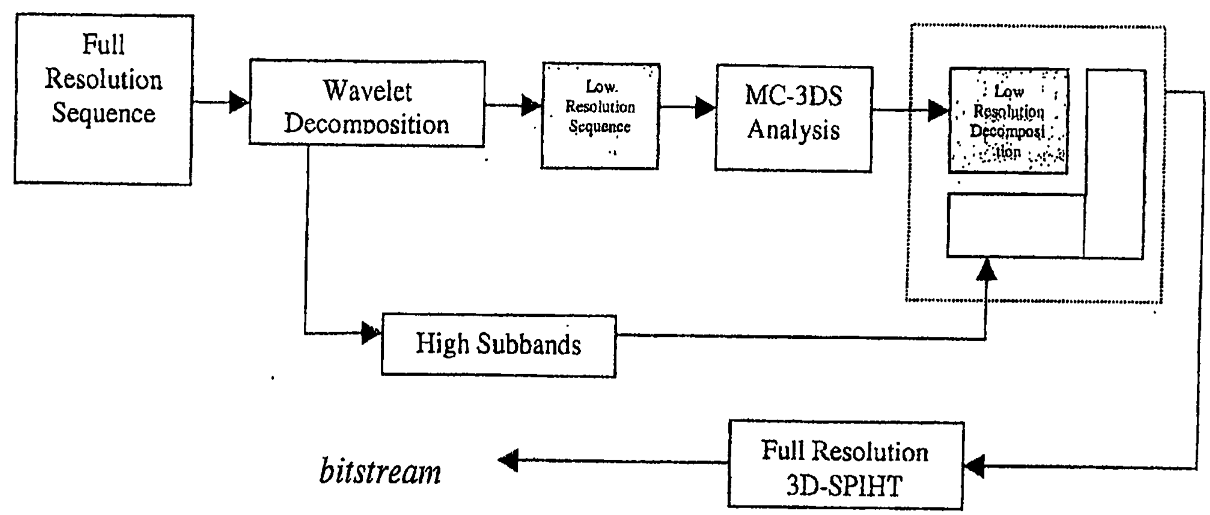
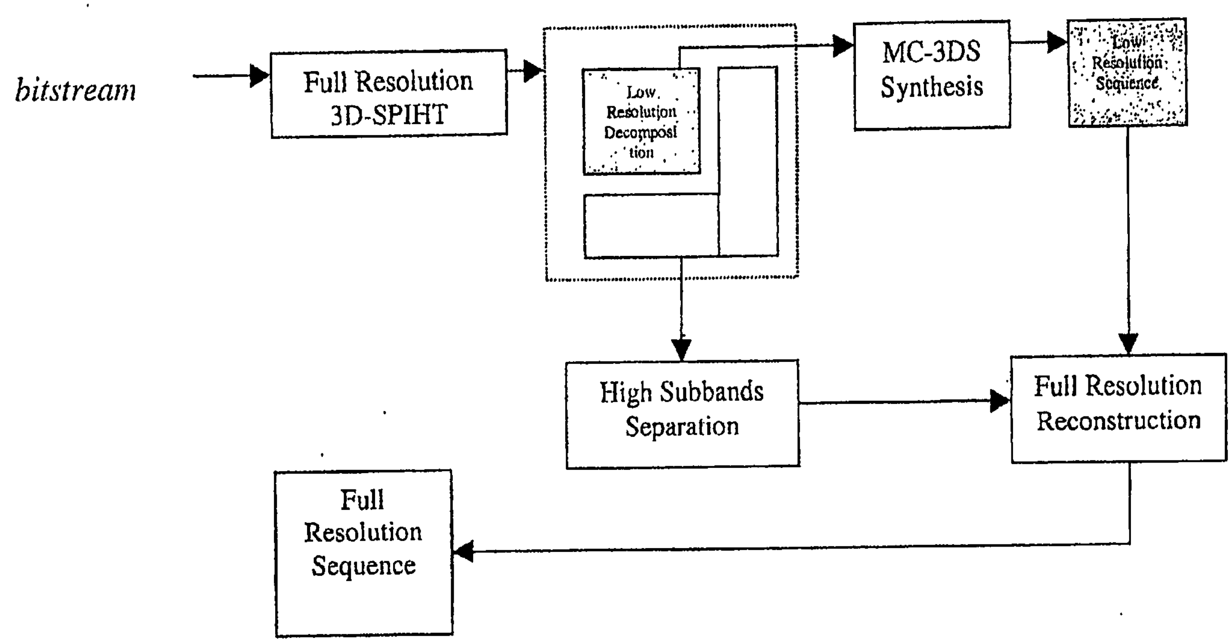
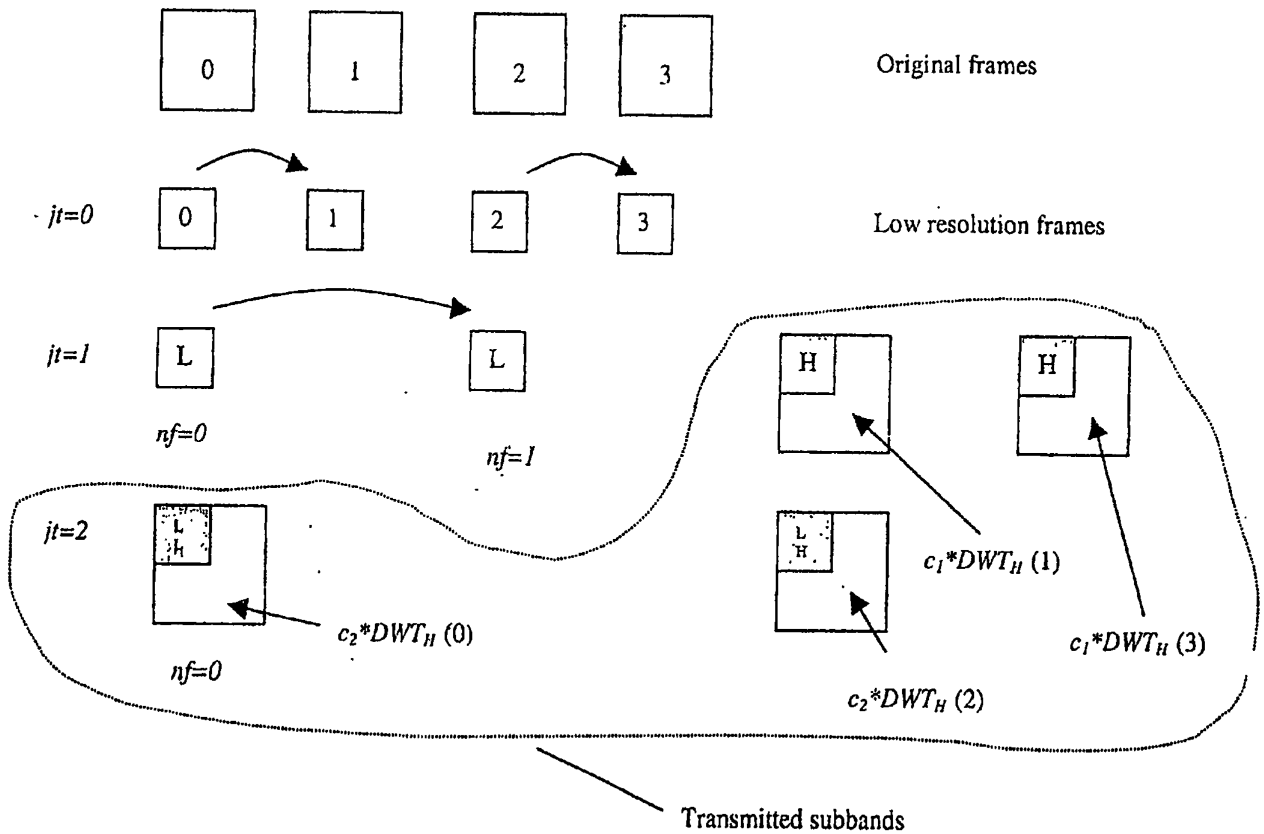

Full Resolution Sequence
Wavelet Decomposition
Low Resolution Sequence
MC-3DS Analysis
Low Resolution Decomposition
High Subbands
Full Resolution 3D-SPIHT
bitstream

FIG.3



bitstream
Full Resolution 3D-SPIHT
Low Resolution Decomposition
MC-3DS Synthesis
Low Resolution Sequence
High Subbands Separation
Full Resolution Reconstruction
Full Resolution Sequence

FIG. 4



0
1
2
3
Original frames
jt=0
0
1
2
3
Low resolution frames
jt=1
L
L
nf=0
nf=1
H
H
jt=2
L
H
L
H
c1*DWTH (1)
c2*DWTH (0)
c1*DWTH (3)
nf=0
c2*DWTH (2)
Transmitted subbands

FIG.5

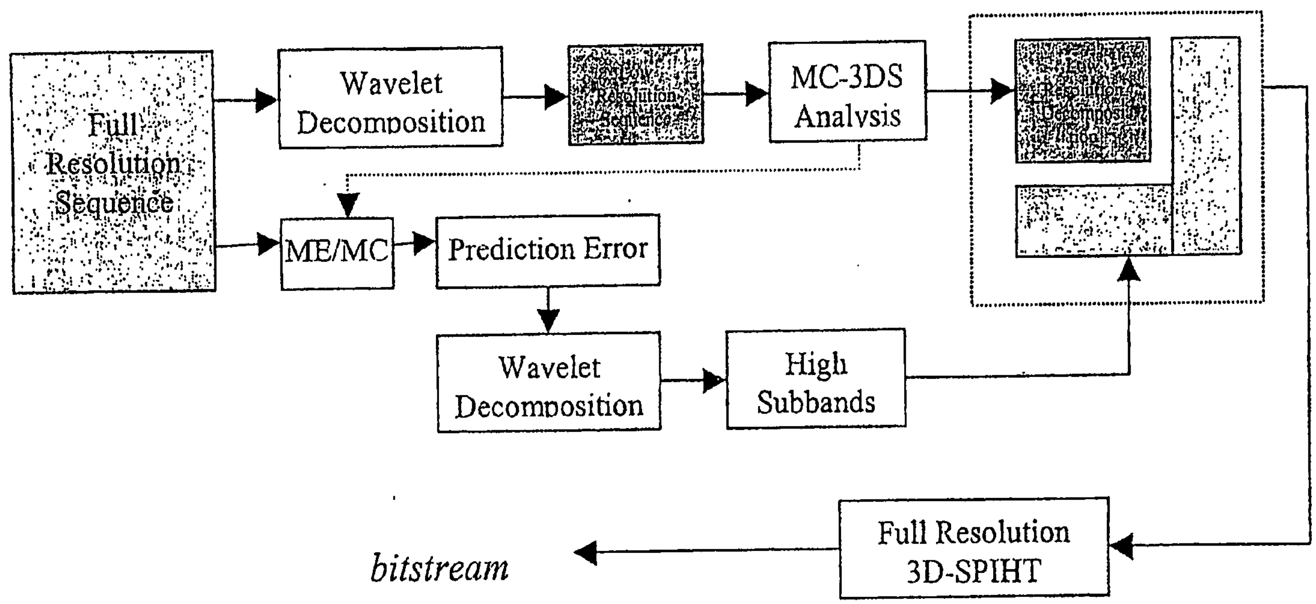

Full Resolution Sequence
Wavelet Decomposition
MC-3DS Analysis
ME/MC
Prediction Error
Wavelet Decomposition
High Subbands
Full Resolution 3D-SPIHT
bitstream

FIG. 6